

## Identification of Disease Genes

This mini-course focuses on the identification of a disease gene using NCBI's human genome assembly. The reference human genome assembly along with integrated maps, literature, and expression information comprises a powerful discovery system for exploring candidate human disease genes.

**Problem:** A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

**We will follow these steps to solve the problem:**

1. Compare an EST to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using MapViewer).
3. Identify whether the EST contains any known SNPs (using dbSNP).
4. Determine whether a mutant form of the gene causes a phenotype (using OMIM).

A web page

(<http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasegene.html>) describes in detail how to perform these steps.

The following handout includes the screen shots of the exercise.

Course developed by : Medha Bhagwat ([bhagwat@ncbi.nlm.nih.gov](mailto:bhagwat@ncbi.nlm.nih.gov) )

## Problem 1:

A laboratory has generated an EST library from a hemochromatosis patient and wants to identify the gene(s) causing the phenotype.

### *Outline:*

We will follow these steps to solve the problem:

1. Compare the EST from a hemochromatosis patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the ESTs and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

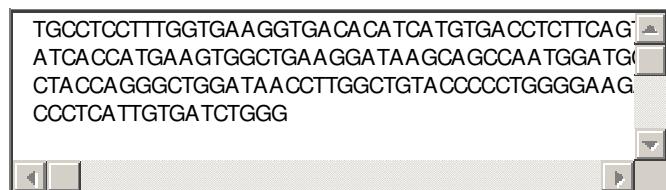
### *Step 1. Compare ESTs to the human genome (using BLAST):*

One way to identify the genes expressing the EST is to compare the EST sequence with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on

#### BLAST (human genome)

Paste the EST sequence provided below in the query box of the BLAST page and start the search by clicking on the “Begin Search” button.

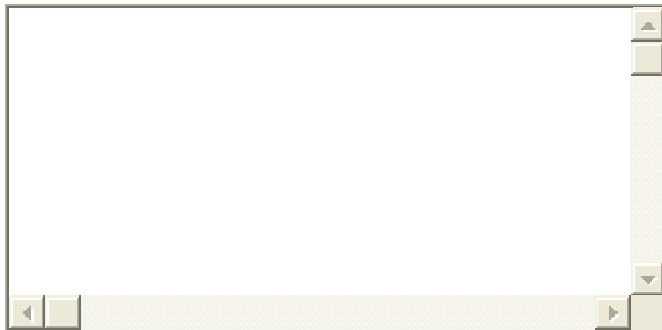
Query EST Sequence:



TGCCTCCCTGGTGAAGGTGACACATCATGTGACCTCTTCAG  
ATCACCATGAACTGGCTGAAAGATAAGCAGCCAATGGATG  
CTACCAGGGCTGGATAACCTTGGCTGTACCCCCTGGGAAG  
CCCTCATTTGATCTGGG

Name the chromosome and the contig that we get as a BLAST hit. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference between the two sequences. Paste your results in the window below.

## Results of BLAST against the human genome

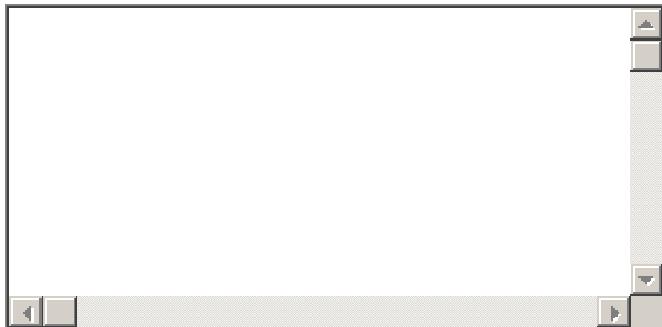


### *Step 2. Identify the gene(s) expressing the ESTs and download their sequences:*

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT\_007592". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene\_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.

The BLAST hit, indicated by the red bar, is in the region of one of the exons of the HFE gene annotated on the human genome. Make the Gene\_seq map a master map by clicking on the arrow at the top of the map. Display the entire HFE gene sequence by clicking on the "dl" link and then on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

### HFE gene sequence



### *Step 3. Determine whether the ESTs contain known SNPs:*

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene\_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and master map) and Gene\_seq. The master map provides detailed information for the map features, in this case SNPs. ". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs1800562. Click on the link for the SNP. There is an A/G SNP is at the nucleotide position 16951392 on the contig NT\_007592 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1? Please note that the SNP results in the Cysteine 282 Tyrosine mutation for the longest protein (expressed by the mRNA NM\_000410) as reported under GeneView.

### *Step 4. Determine whether the mutant HFE gene causes a phenotype:*

Go back to the Map Viewer report. Make the Gene\_seq map as the master map. Select the link to the OMIM database. It takes us to the OMIM report for the HFE gene that details how mutations in the HFE gene are associated with a phenotype, hemochromatosis. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Cys282Tyr variant mentioned in the list? Which phenotype does it cause?

## Summary:

This mini-course describes steps to identify the gene expressing an EST obtained from a hemochromatosis patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

Step 1: The query EST sequence was found to align contig NT\_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).

Step 2: The query EST was found to be expressed by the HFE gene.

Step 3: The query EST sequence contains a known SNP (G/A with respect to nucleotide 16951392 on contig NT\_007592.14).

Step 4: Mutations in the HFE gene are associated with hemochromatosis.

# Snapshots

*Step 1: Compare EST against the human genome.*

The screenshot shows the NCBI homepage. At the top, there's a navigation bar with links for PubMed, Entrez, BLAST (which is highlighted with a pink arrow), OMIM, Books, TaxBrowser, and Structure. Below the navigation bar is a search bar with the placeholder "Search Entrez for" and a "Go" button. To the left of the main content area is a sidebar with links to Site Map, Guide to NCBI resources, About NCBI (with a sub-link to "An introduction for researchers, educators and the public"), GenBank (with a sub-link to "Sequence submission support and software"), Literature databases (with a sub-link to "PubMed, OMIM, Books, and PubMed Central"), Molecular databases (with a sub-link to "Sequences, structures, and taxonomy"), and Genomic. The main content area features a section titled "What does NCBI do?" which describes the organization's mission and history. It also highlights "PubMed Central" as "An archive of life sciences journals" with "Free fulltext", "Over 100,000 articles from over 130 journals", and "Linked to PubMed and fully searchable". A note states that access requires no registration or fee. Another section, "Entrez Gene", explains how users can search for gene information across various sources. To the right of these sections is a "Hot Spots" sidebar containing a list of research areas: Clusters of orthologous groups, Electronic PCR, E-Utilities, Gene expression omnibus, Genes and disease, Human genome resources, Human-mouse homology maps, LocusLink, Malaria genetics & genomics, Map Viewer, and MHC.

NCBI → BLAST		Latest news: 6 December 2005 : BLAST 2.2.13 released	
<b>About</b> <ul style="list-style-type: none"> <li>Getting started</li> <li>News</li> <li>FAQs</li> </ul> <b>More info</b> <ul style="list-style-type: none"> <li>NAR 2004</li> <li>NCBI Handbook</li> <li>The Statistics of Sequence Similarity Scores</li> </ul> <b>Software</b> <ul style="list-style-type: none"> <li>Downloads</li> <li>Developer info</li> </ul> <b>Other resources</b> <ul style="list-style-type: none"> <li>References</li> <li>NCBI Contributors</li> <li>Mailing list</li> <li>Contact us</li> </ul>		<p><b>The Basic Local Alignment Search Tool (BLAST)</b> finds regions of local similarity between sequences. The program compares nucleotide or protein sequences to sequence databases and calculates the statistical significance of matches. BLAST can be used to infer functional and evolutionary relationships between sequences as well as help identify members of gene families.</p>	
<b>Nucleotide</b> <ul style="list-style-type: none"> <li>Quickly search for highly similar sequences (megablast)</li> <li>Quickly search for divergent sequences (discontiguous megablast)</li> <li>Nucleotide-nucleotide BLAST (blastn)</li> <li>Search for short, nearly exact matches</li> <li>Search trace archives with megablast or discontiguous megablast</li> </ul>		<b>Protein</b> <ul style="list-style-type: none"> <li>Protein-protein BLAST (blastp)</li> <li>Position-specific iterated and pattern-hit initiated BLAST (PSI- and PHI-BLAST)</li> <li>Search for short, nearly exact matches</li> <li>Search the conserved domain database (rpsblast)</li> <li>Protein homology by domain architecture (cdart)</li> </ul>	
<b>Translated</b> <ul style="list-style-type: none"> <li>Translated query vs. protein database (blastx)</li> <li>Protein query vs. translated database (tblastn)</li> <li>Translated query vs. translated database (tblastx)</li> </ul>		<b>Genomes</b> <ul style="list-style-type: none"> <li>Human, mouse, rat, chimp, cow, pig, dog, sheep, cat</li> <li>Chicken, puffer fish, zebrafish</li> <li>Fly, honey bee, other insects</li> <li>Microbes, environmental samples</li> <li>Plants, nematodes</li> <li>Fungi, protozoa, other eukaryotes</li> </ul>	
<b>Special</b> <ul style="list-style-type: none"> <li>Search for gene expression data (GEO BLAST)</li> <li>Align two sequences (bl2seq)</li> </ul>		<b>Meta</b> <ul style="list-style-type: none"> <li>Retrieve results</li> </ul>	

BLAST Human Sequences. - Microsoft Internet Explorer

File Edit View Favorites Tools Help

Back Favorites Go Links

Address: http://www.ncbi.nlm.nih.gov/genome/seq/BlastGen/BlastGen.cgi?taxid=9606

NCBI Home > Genomic Biology > Human Genome Resources > BLAST

Search Map Viewer Go Clear

**BLAST Human Sequences.**

Enter an accession, gi, or a sequence in FASTA format:  
TGCTCTTTGGTGAAGGTGACACATCATGTGACCTTCAGTGACCCTACGGTGTGGGCC  
TTGAACTACTACCCCCAGAAC  
ATCACCAGTGAAGTGGCTGAAGGATAAGCAGCCAATGGATGCCAAGGAGTTGAAACCTAAAGACGT  
ATTGCCCAATGGGGATGGGAC  
CTACCAAGGGCTGGATAACCTTGCTGTACCCCTGGGGAAAGAGCAGAGATAACGTACCCAGGTGG  
AGCACCCAGGCCTGGATCAGC

Or, choose a file to upload

**Database:**

genome (all assemblies)

genome (all assemblies)

genome (reference only)

HTGS

RefSeq RNA

RefSeq protein

Non-RefSeq RNA

Non-RefSeq protein

Build RNA

Build protein

Ab initio RNA

Ab initio protein

Advanced options:

Alignments

Begin Search

Local intranet

**NCBI** **formatting BLAST**

Nucleotide Protein Translations Retrieve results for an RID

Your request has been successfully submitted and put into the Blast Queue.

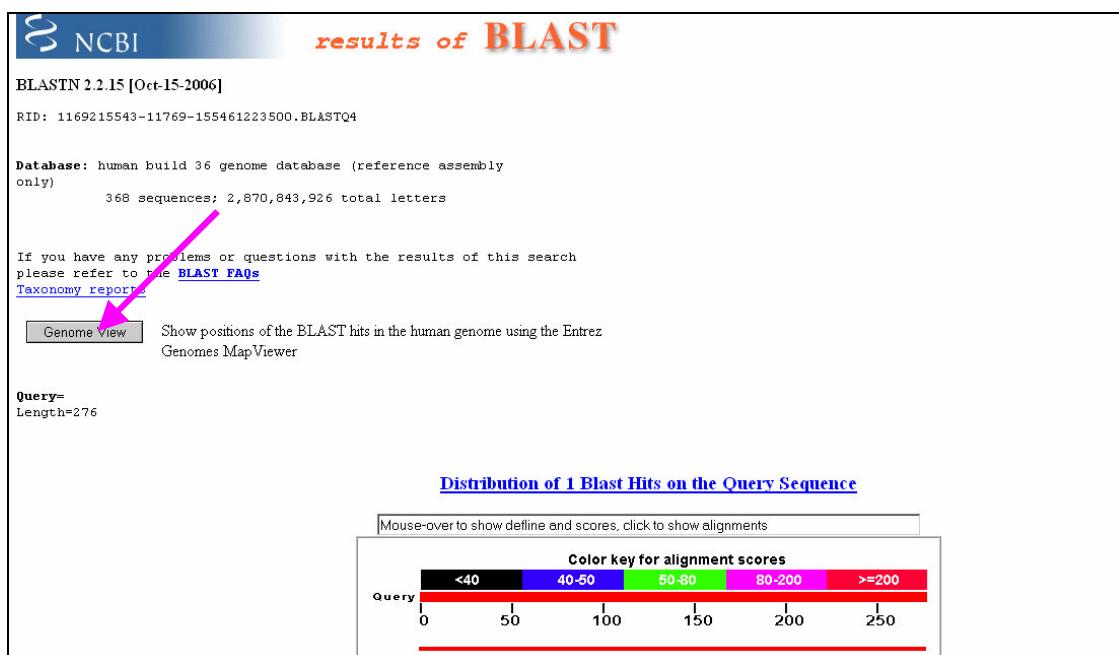
Query = (276 letters)

The request ID is 1169215543-11769-155461223500.BLASTQ4

**Format!** or **Reset all**

The results are estimated to be ready in 17 seconds but may be done sooner.

Please press "FORMAT!" when you wish to check your results. You may change the formatting options for your result via the form below and press "FORMAT!" again. You may also request results of any other valid request ID to see other recent jobs.



Alignments

>[ref|NT\\_007592.14|Hs6\\_7749](#) D Homo sapiens chromosome 6 genomic contig  
Length=48945890

Features in this part of subject sequence:  
[hemochromatosis protein isoform 11 precursor](#)  
[hemochromatosis protein isoform 10 precursor](#)

Score = 505 bits (273), Expect = 6e-141  
 Identities = 275/276 (99%), Gaps = 0/276 (0%)  
 Strand=Plus/Plus

Query 1	TGCCTCCTTGGTGAAGGTGACACATCATGTGACCTCTCAGTGACCACACTACGGTGTC	60
Sbjct 16951164	TGCCTCCTTGGTGAAGGTGACACATCATGTGACCTCTCAGTGACCACACTACGGTGTC	16951223
Query 61	GGGCCTTGAACTACTACCCCCAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCAA	120
Sbjct 16951224	GGGCCTTGAACTACTACCCCCAGAACATCACCATGAAGTGGCTGAAGGATAAGCAGCAA	16951283
Query 121	TGGATGCCAAGGAGTTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAAGG	180
Sbjct 16951284	TGGATGCCAAGGAGTTCGAACCTAAAGACGTATTGCCCAATGGGGATGGGACCTACCAAGG	16951343
Query 181	GCTGGATAACCTTGGCTGTACCCCTGGGGAGAGCAGAGATATCGTACCAAGTGGAGC	240
Sbjct 16951344	GCTGGATAACCTTGGCTGTACCCCTGGGGAGAGCAGAGATATCGTCCAGTGGAGC	16951403
Query 241	ACCCAGGCCCTGGATCAGCCCTCATTGTGATCTGGG	276
Sbjct 16951404	ACCCAGGCCCTGGATCAGCCCTCATTGTGATCTGGG	16951439

Result: The EST sequence is aligned to the contig NT\_007592.14 on chromosome 6 with one nucleotide difference (G to A with respect to the nucleotide 16951392 on the contig).

*Step 2: Identify the gene(s) expressing the EST and download their sequences*

 NCBI      results of BLAST

BLASTN 2.2.15 [Oct-15-2006]

RID: 1164596270-6327-66313738572.BLASTQ2

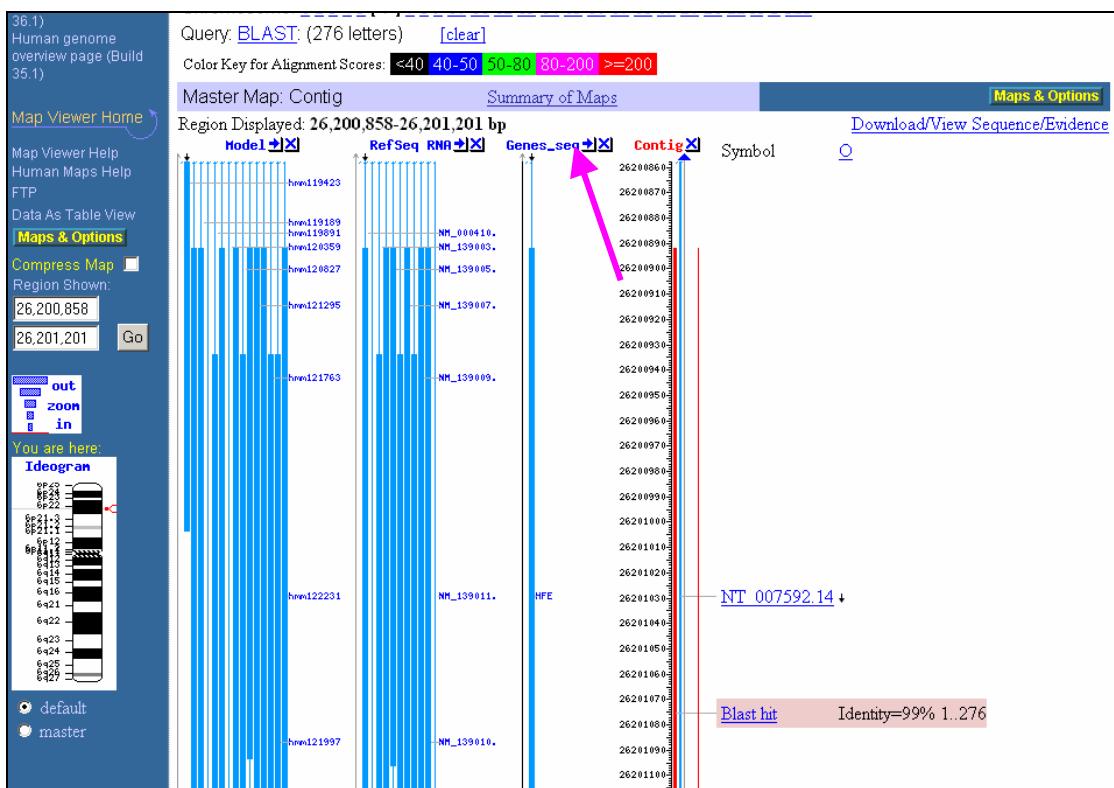
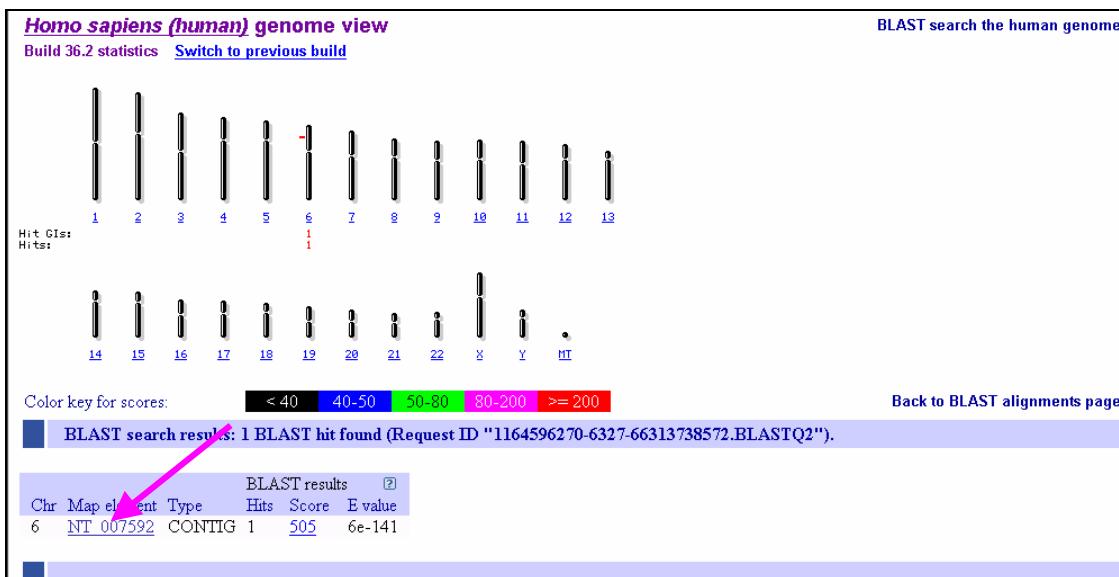
**Database:** human build 36 genome database (reference assembly only)  
 368 sequences; 2,870,843,926 total letters

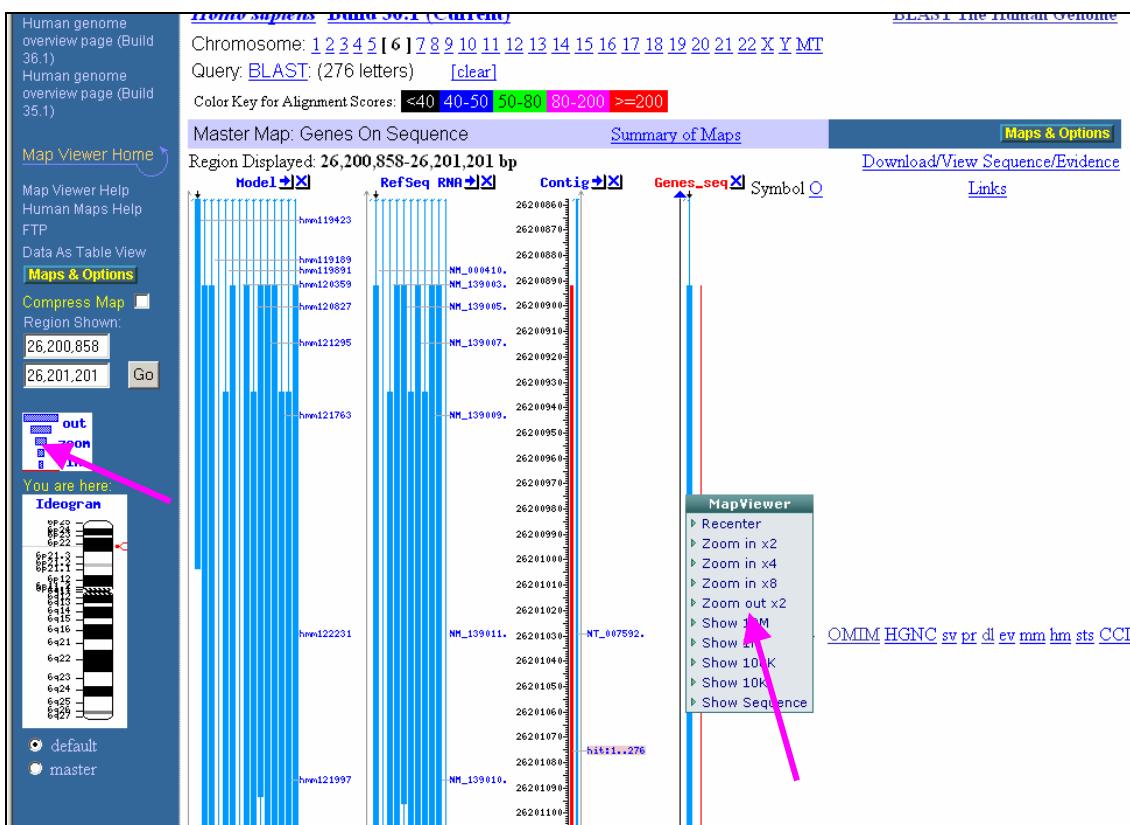
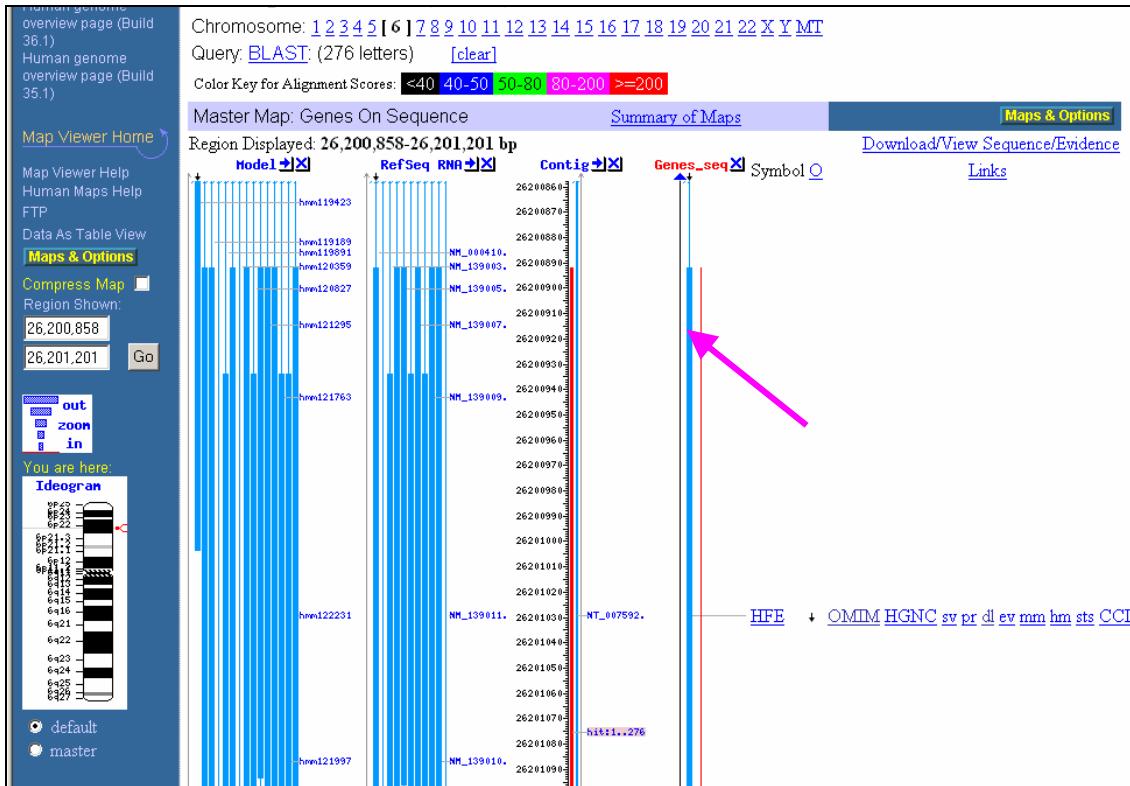
If you have any problems or questions with the results of this search  
 please refer to the [BLAST FAQs](#)  
[Taxonomy reports](#)

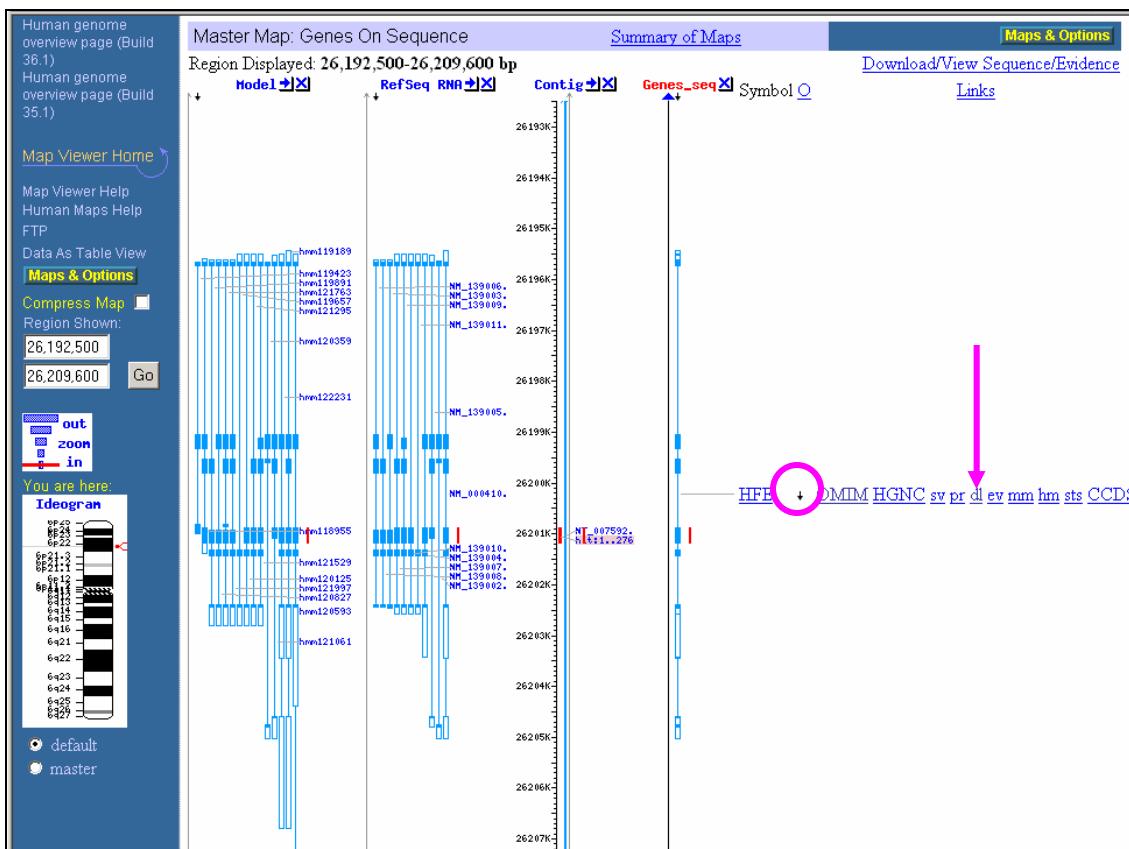
 **Genome View** Show positions of the BLAST hits in the human genome using the Entrez Genomes MapViewer

**Query=**  
 Length=276

[Distribution of 1 Blast Hits on the Query Sequence](#)







*homo sapiens* (Build 36.1)  
Region to retrieve (in chromosome coordinates):

Chromosome:  Strand:

From:  adjust by:   
to:  adjust by:  Change Region/Strand

Sequence Format:

This chromosome region corresponds to the contig region(s):

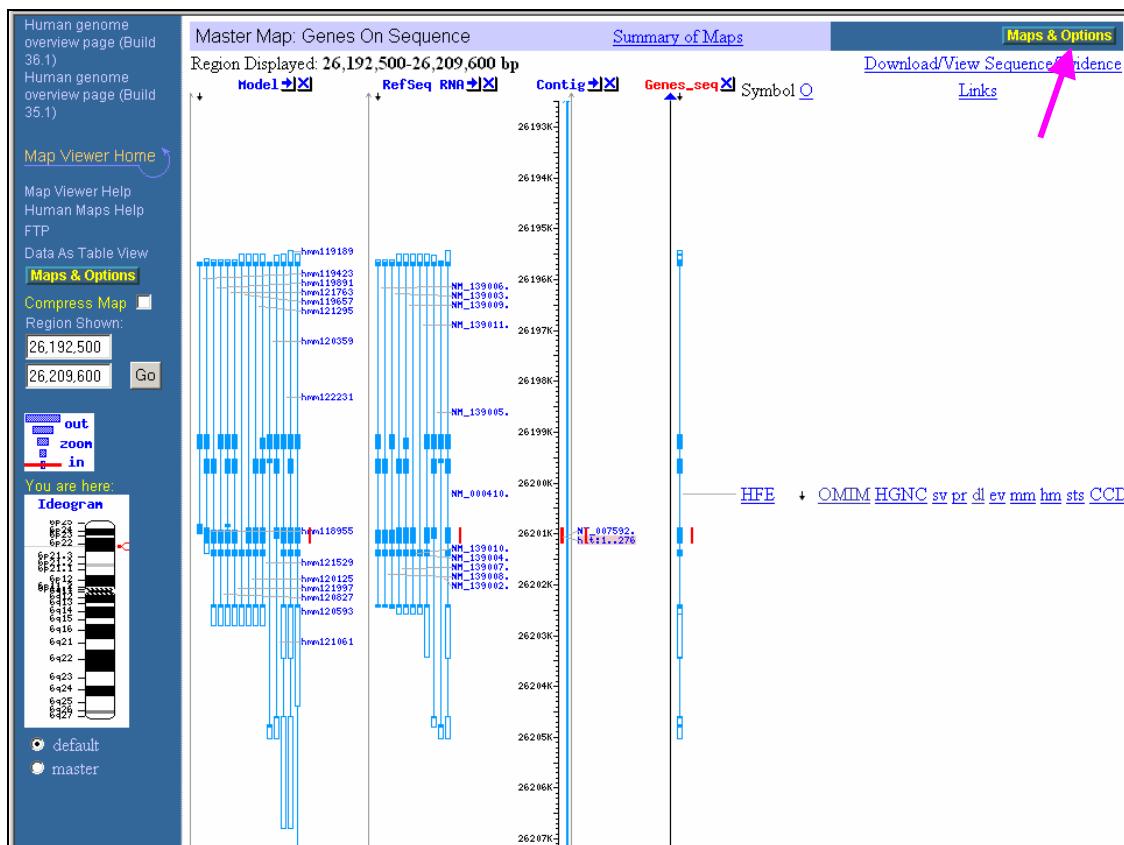
Contig	start	stop	strand
NT_007592.14	16945699	16955310	+ <a href="#">Display</a> <a href="#">Save to Disk</a> <a href="#">View Evidence</a> <a href="#">ModelMaker</a>

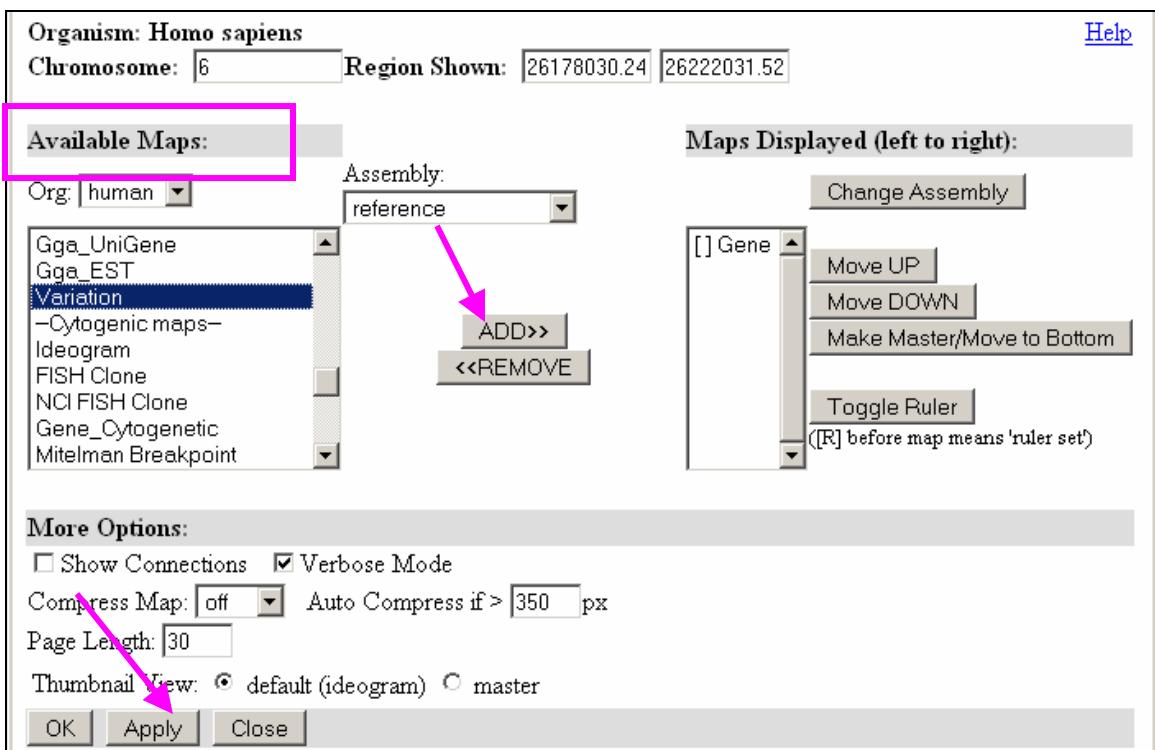
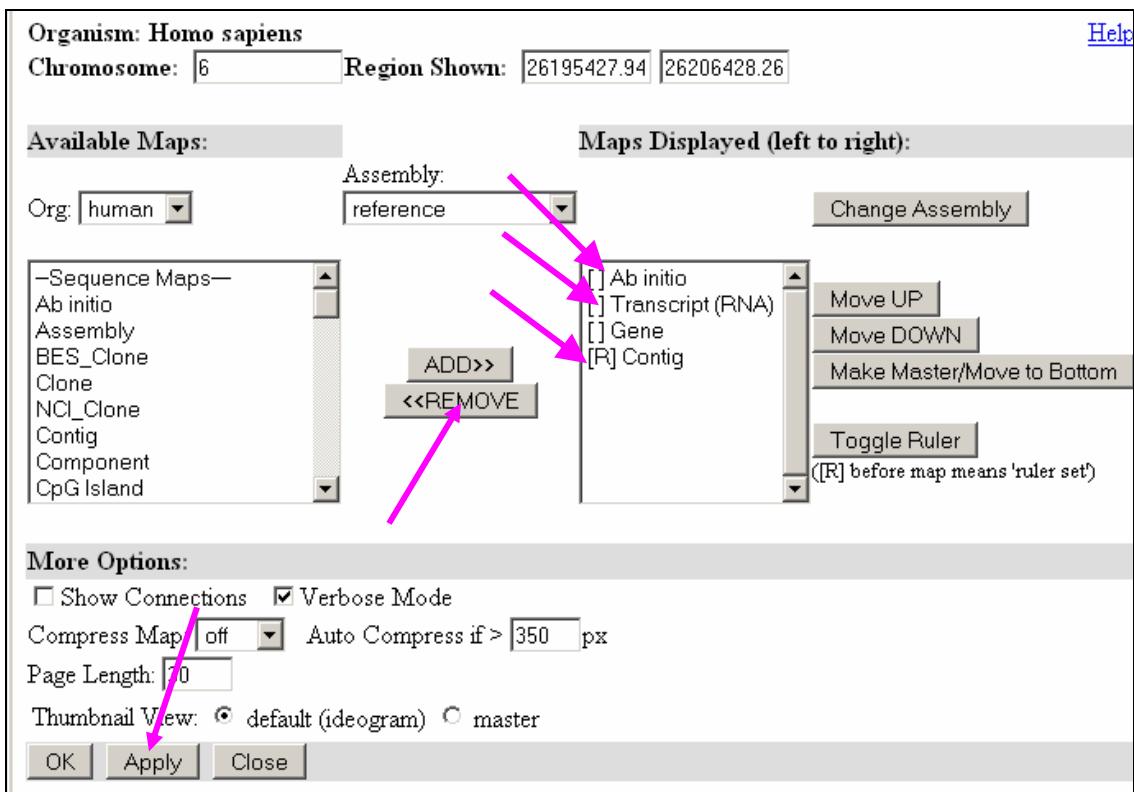
Range: from 16945699 to 16955310 Show whole sequence Reverse complemented strand Refresh  
□ 1: NT\_007592. Reports Homo sapiens chro...[gi:51465675]

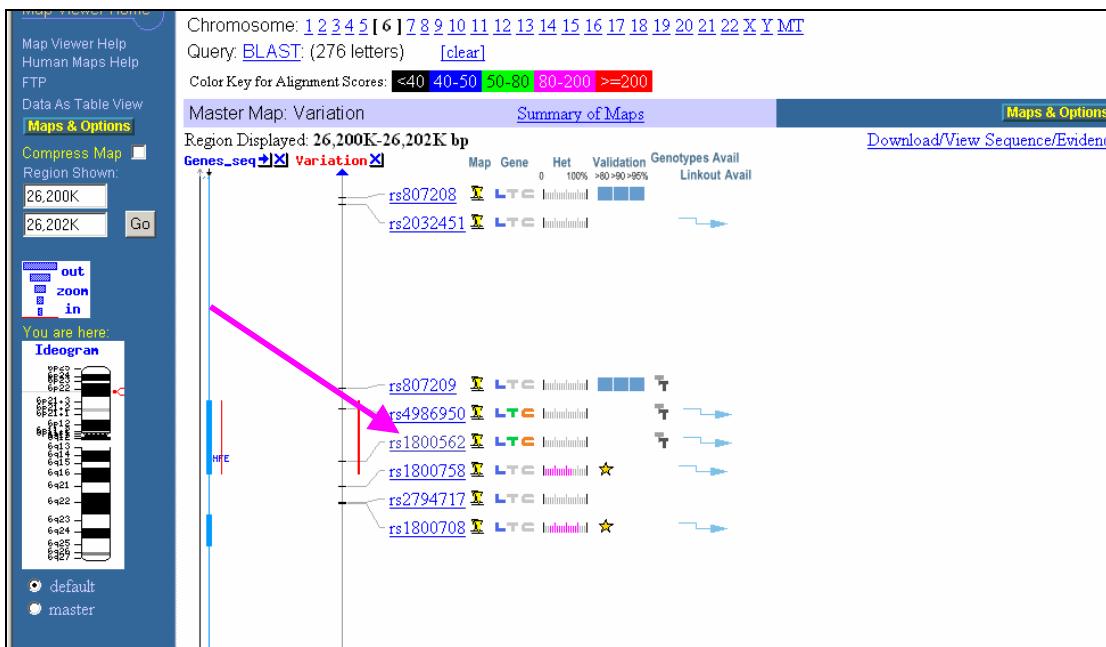
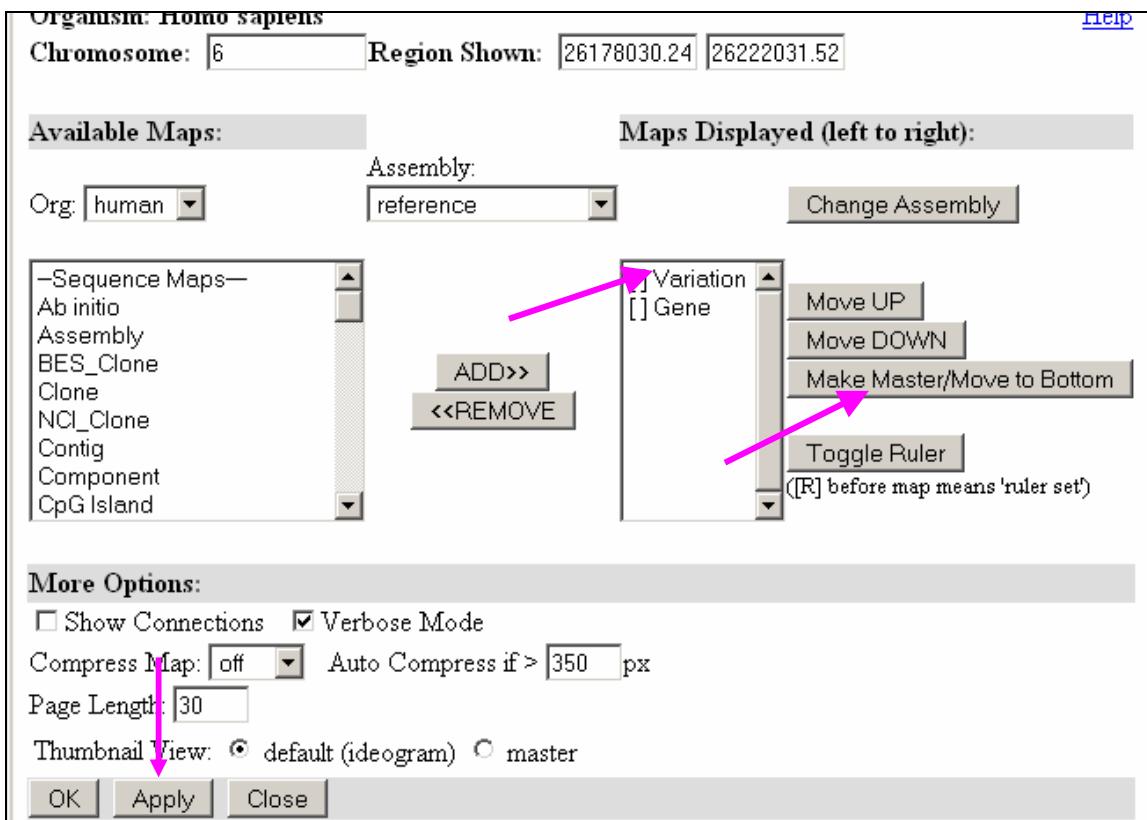
>ref|NT\_007592.14|Hs6\_7749:16945699-16955310 Homo sapiens chromosome 6 genomic contig  
GGGGACATGGATCACCTAGTGTTCACAAGCAGGTACCCCTCTGCTGTAGGAGAGAGAACTAAAGTT  
TGAAAGACCTGTGCTTTCACCAAGGAATTTACTGGCATCTCCCTGAGCCTAGGCATAATAGCTGTAGGG  
TGACTCTGGAGGCCATCCCCCTTCCCCCCCCAAAAGAACGGGAGATTAAACGGGGACGTGCGGGCCA  
GAGCTGGGAAATGGGCCCGGAGGCCAGGCCGGCGCTTCCTCTGATGCTTTGCAGACCGCGGTCT  
GCAGGGGGCCTGCTGCGTAGTCGAGGGCTGCGGGCAACTAGGGCGGGCGGGGGTGGAAAAATCG  
AAACTAGTTTTCTTTCGCTGGAGTTGCTAACTTGGAGGACCTGCTCAACCTATCCGAAGCC  
CCTCTCCCTACTTTCTGCTGAGGCCAGACCCGGTAGGGGAGTGCCTACCACTGAACTGCAAGATAGGGGTCCT  
CGCCCCCAGGACCTGGCCCCCTCCCCGGCTGCTGGGGCTGAGCTGGAGGTGACTTTGGAACCGGCCACTCCC  
TTCCCCCAACTAGAATGCTTTAAATAAACTCTGCTAGTGTCTACTTGAACGCTGAGCTAAACCTGGGCTC  
CTTGAACTGGAACTCGGTTTATTCAAATGTCAAGCTGTGCAAGTTTTCCCCAGTCATCTCAAACAG  
GAAGTTCTCCCTGAGTGTGCGAGAAGGGCTGAGCAACCCACAGCAGGGATCCGACGGGTTTCCAC  
CTCAGAACGAATGCGTTGGCGGTGGGGCGCAGAAGAGTGGCTTGGGATCTGAATTCTCACCAATT  
CACCCACTTTGGTAGAGACCTGGGGTGGAGGTCTAGGGTGGGGCTCTGAGAGAGGCCTACCTCGG  
GCCTTCCCCACTCTGGCAATTGTTCTTGTGGAAATAAAGTATATGTTAGTTGAACGTTGA  
ACTGAACAAATTCTCTTTCGGCTAGGCTTATTGATTTCAAATGTGCTGTAAATAAGGCTCTCA  
CAAAGTACTGATAATGAACATGTAAGCAATGCACTCACTTCAAGTTACATTACATATGATCTTATTG  
ATTTTCACTAGGCATAGGGAGGTAGGGAGCTAATAATACGTTTATTTACTAGAAGTTAACTGGAATTCA  
ATTATAAACTCTTCACTGGTACAAAGAACATAAAATAATCTGTTTCTGATGTTATTTCAGTACTAC  
AGCTGCTCTAACTCTTAGGTGACAGTGTAGTTGGCTGTAGTGCAGCTGGTCTGTGGGTCAACACGC  
CGGGCTGAGCACACGACTTGTAGTTGGTACTACGTGTATCCACATTACATGACAAAATGAGGC  
ATGGCACGGCTGCTTCTGGCAATTATTCAATGGTACACTGGCTTGGTGGCAGAGCTCATGTCT

Result: The query EST is expressed by the HFE gene.

*Step 3: Determine whether the EST contains any known SNPs*







**NCBI** Single Nucleotide Polymorphism

PubMed Nucleotide Protein Genome Structure PopSet Taxonomy OMIM Books SNP

Search Entrez [SNP] for

**BUILD 126**

**GENERAL**  
 Contact Us  
 dbSNP Homepage  
 SNP Science Primer  
 Announcements  
 dbSNP Summary  
 FTP Download

refSNP ID: rs1800562	Allele	Links
Organism: human ( <i>Homo sapiens</i> )	SNP:	
Molecule Type: Genomic	Variation Class:	single nucleotide polymorphism
Created/Updated in build: 89/123	Alleles:	A/G
Map to Genome Build: <a href="#">36.1</a>	Ancestral Allele:	Not available

**SEARCH**  
 Entrez SNP  
 Blast SNP  
 Batch Query  
 By Submitter  
 New Batches  
 Method  
 Population  
 Detail  
 Class  
 Publication  
 Chromosome Report  
 Locus Information  
 STS Markers  
 Free Form Search

**Fasta sequence (Legend)**

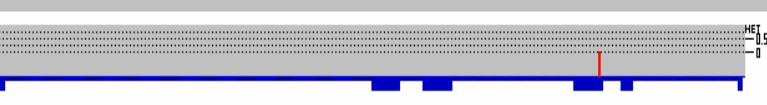
```
>gnl|dbSNP|rs1800562|allelePos=202|totalLen=450|taxid=9606|snpclass=1|alleles=A/G|mol=Genomic|build=113
ATGTGAYCTC TTCAGTGACC ACTCTACGGT GTCGGGCCTT GAACTACTAC CCCCAGAAC
TCACCATGAA GTGGCTGAAG GATAAAGCAGC CAATGGATGC CAAGGAGTTC GAACTCTAAAG
ACGTATTGCC CAATGGGAT GGGACCTACC AGGGCTGGAT AACCTTGCT GTACCCCCCTG
GGGAAGAGCA GAGATATACG T
R
CCAGGTGGAG CACCCAGGCC TGGATCAGCC CCTCATTTG TGATCTGGGTA TGTGACTGAT
GAGGCCAGG AGCTGAGAAA ATCTATTG GGTRAGAGG AGTGCCTGAG GAGGTAATTA
TGGCACTGAG ATGAGGATCT GCTCTTGT AGGGGGTGG CTGAGGGTGG CAATCAAAGG
CTTAACTTG CTTTTCTGT TTTAGAGCCC TCACCGTCTG GCACCCTAGT CATTGGAGTC
ATCAGTGG
```

**Integrated Maps:**

NCBI Map Viewer: rs1800562 maps exactly once on NCBI human [chromosome 6](#)

Chromosome	Contig accession	Contig position	Chromosome	Hit orientation	Contig	Assembly	Group label	Contig label	Neighbor	SNP flank position
6	<a href="#">NT_007592.14</a>	1691392	26201120	plus	G	ref_assembly	reference	<a href="#">view</a>	201	
6	<a href="#">NW_922984.1</a>	25717681	27322425	plus	G	alt_assembly_1	Celera	<a href="#">view</a>	201	

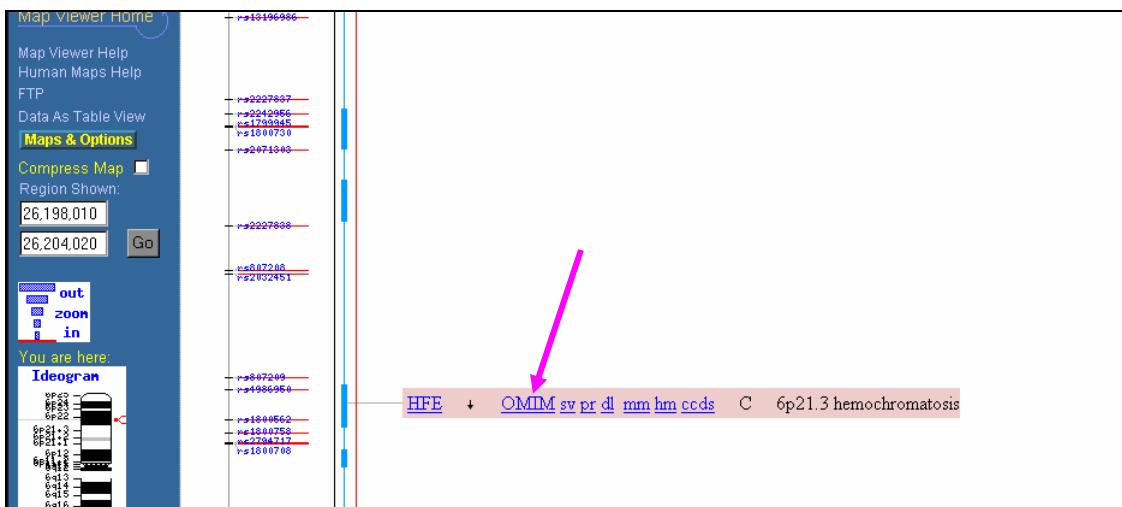
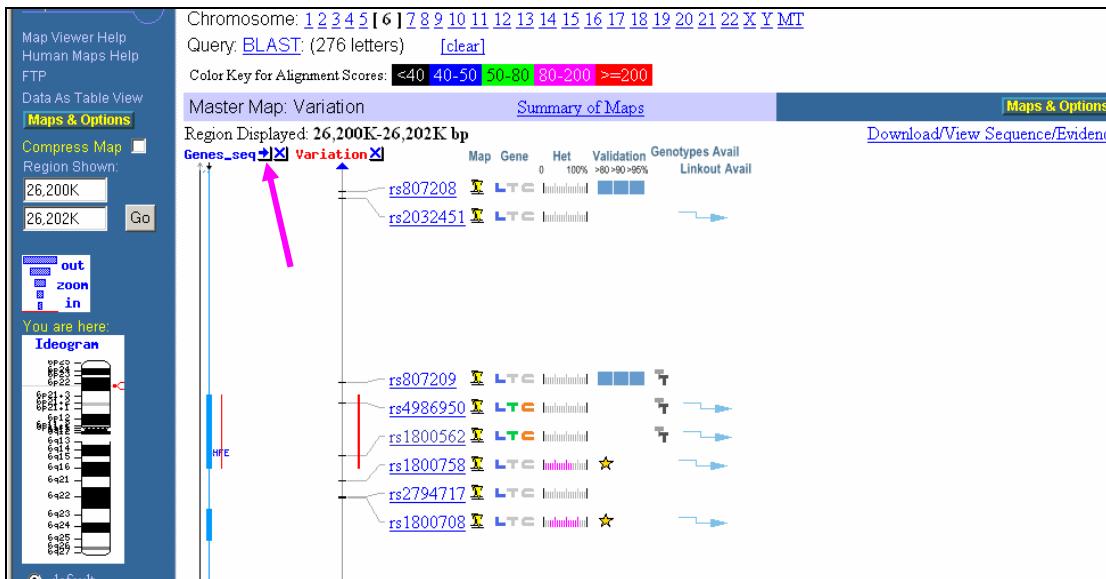
Click to see [all](#) [details](#) [this frequency](#) [genetic map](#) [haplotypes](#) [variations associated with this gene](#).

Group Label	Contig->mRNA	Gene Model (contig mRNA transcript) <a href="#">Color Legend</a>
reference <a href="#">NT_007592-&gt;NM_000410</a>	<a href="#">sv function</a>	
Contig	16951392	16951392

Group label	Contig-->mRNA-->Protein	Contig position	mRNA orientation	mRNA pos	Function	dbSNP allele	Protein residue pos	Codon pos	Amino acid pos
reference <a href="#">NT_007592-&gt;NM_000410-&gt;NP_000401</a>	16951392	forward	1066	nonsynonymous	A	Tyr [Y]	2	282	
				contig reference	G	Cys [C]	2	282	

Result: The EST sequence contains a known SNP (G/A with respect to the nucleotide 16951392 on contig NT\_007592.14).

## Step 4: Determine whether a mutant HFE gene causes a phenotype



**NCBI**

MIM +235200

Description  
Clinical Features  
Other Features  
Inheritance  
Mapping  
Heterogeneity  
Molecular Genetics  
Genotype/Phenotype  
Correlations  
Diagnosis  
Clinical Management  
Population Genetics  
Pathogenesis  
Cloning  
Biochemical Features  
Gene Structure  
Gene Function  
Nomenclature  
Animal Model  
History  
Allelic Variants  
• View List  
See Also  
References  
Contributors  
Creation Date  
Edit History

• Clinical Synopsis  
• Gene map

**+235200**  
**HEMOCHROMATOSIS; HFE**

GeneTests, Links

**Alternative titles: symbols**

**HLAH**  
**HEMOCHROMATOSIS, HEREDITARY; HH**  
**HFE GENE, INCLUDED; HFE, INCLUDED**

Gene map locus [6p21.3](#)

**TEXT**

**DESCRIPTION**

The clinical features of hemochromatosis include cirrhosis of the liver, diabetes, hypermelanotic pigmentation of the skin, and heart failure. Primary hepatocellular carcinoma (HCC; [114550](#)), complicating cirrhosis, is responsible for about one-third of deaths in affected homozygotes. Since hemochromatosis is a relatively easily treated disorder if diagnosed, this is a form of preventable cancer.

**NCBI**

MIM +235200

Description  
Clinical Features  
Other Features  
Inheritance  
Mapping  
Heterogeneity  
Molecular Genetics  
Genotype/Phenotype  
Correlations  
Diagnosis  
Clinical Management  
Population Genetics  
Pathogenesis  
Cloning  
Biochemical Features  
Gene Structure  
Gene Function  
Nomenclature  
Animal Model  
History  
Allelic Variants  
• View List  
See Also  
References  
Contributors  
Creation Date  
Edit History

• Clinical Synopsis  
• Gene map

**+235200**  
**HEMOCHROMATOSIS; HFE**

GeneTests, Links

**ALLELIC VARIANTS**  
(selected examples)

- [0001 HEMOCHROMATOSIS](#) [HFE, CYS282TYR]
- [0002 HEMOCHROMATOSIS](#) [HFE, HIS62ASP]
- [0003 HEMOCHROMATOSIS](#) [HFE, SER65CYS]
- [0004 HFE INTRONIC POLYMORPHISM](#) [HFE, 5569G-A]
- [0005 HFE POLYMORPHISM](#) [HFE, VAL53MET]
- [0006 HFE POLYMORPHISM](#) [HFE, VAL59MET]
- [0007 PORPHYRIA VARIEGATA](#) [HFE, GLN127HIS]
- [0008 HEMOCHROMATOSIS](#) [HFE, ARG330MET]
- [0009 HEMOCHROMATOSIS](#) [HFE, ILE105THR]
- [0010 HEMOCHROMATOSIS](#) [HFE, GLY93ARG]
- [0011 HEMOCHROMATOSIS](#) [HFE, GLN283PRO]

Result: Mutations in the HFE gene are associated with hemochromatosis disease.

## Problem 2:

<http://www.ncbi.nlm.nih.gov/Class/minicourses/diseasegene2.html>

A laboratory has generated an EST library from a sickle cell anemia patient and wants to identify the gene(s) causing the phenotype. Sick cell anemia is a disease in which the red blood cells are curved in shape, and which causes pain and fever.

### *Outline:*

We will follow these steps to solve the problem:

1. Compare an EST from a sickle cell anemia patient to the human genome (using BLAST).
2. Identify the gene(s) aligning with the EST and download their sequences (using Map Viewer).
3. Identify whether the EST contains any known nucleotide variations (single nucleotide polymorphisms) (using dbSNP).
4. Determine whether a mutant form of the gene is known to cause a phenotype (using OMIM).

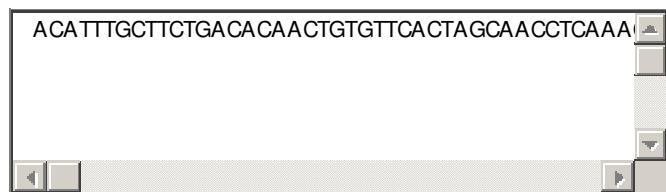
### *Step 1. Compare ESTs to the human genome (using BLAST):*

One way to identify the genes expressing the ESTs is to compare the EST sequence with the human genome assembly and the genes annotated on it. To access the specialized BLAST page for searching against the human genome assembly, click on

#### BLAST (human genome)

Paste the EST sequence provided below in the query box of the BLAST page and start the search by clicking on the “Begin Search” button.

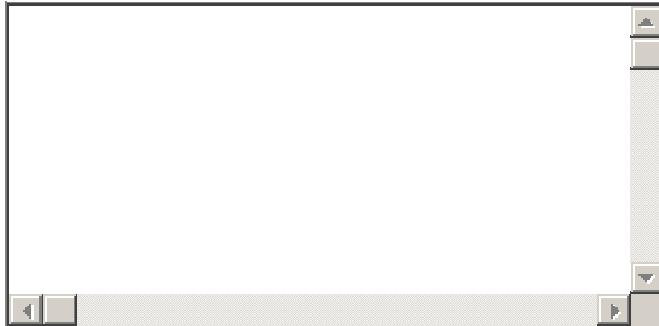
Query EST Sequence:



The image shows a screenshot of a computer screen displaying a BLAST search interface. In the center is a large text input field containing the sequence: ACATTTGCTTCTGACACAAGTGTTCACTAGCAACCTCAA. Below this input field are several small, semi-transparent rectangular boxes, likely indicating where specific nucleotides are being highlighted or aligned. At the bottom of the input field are standard scroll bar controls (up and down arrows, a scroll wheel, and horizontal scroll bars). The background of the interface is white, and the overall appearance is that of a standard web-based search tool.

Name the chromosome and the contig that we get as a BLAST hit. Note that the similarity is on the minus strand of genome. Is the EST sequence 100% identical to the genomic sequence? Note the nucleotide difference between the two sequences. Paste your results in the window below.

## Results of BLAST against the human genome

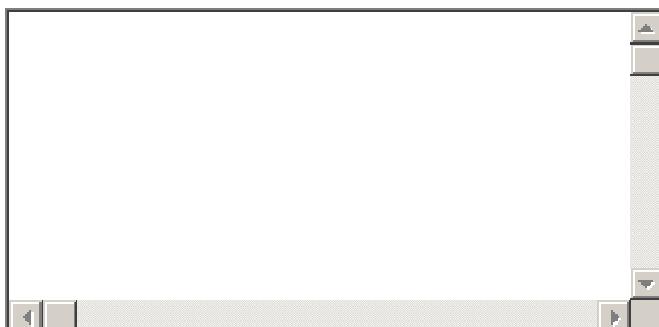


### *Step 2. Identify the gene(s) expressing the EST and download their sequences:*

To visualize the BLAST hit on the genome using Map Viewer, click on the "Genome View" button at the top of the results page, then on the Map element "NT\_009237". Currently, 4 maps should be displayed (Contig, Model, RNA and Gene\_seq). Zoom out 2 or 4 times by clicking on right most contig map and selecting the appropriate option.

The best BLAST hits, indicated by the red bars, are in the region of two exons of the HBB gene annotated on the human genome. Make the Gene\_seq map a master map by clicking on the arrow at the top of the map. Note that the gene is annotated on the minus strand. To display the entire HBB gene sequence, click on the "dl" link, choose minus strand from the pull down menu, click on "Change Region/Strand" and display the sequence by clicking on "Display". Copy the sequence and paste it in the area provided below. We will use it later to obtain the exon-intron structure. You can adjust the nucleotide locations to download the upstream or downstream sequence by using the "adjust by" and "Change Region/Strand" option.

### HBB gene sequence



### *Step 3. Determine whether the EST contains known SNPs:*

Go back to the Map Viewer report. Click on the Maps and Options link. Remove all the maps except the Gene\_seq map by selecting the map under the Maps Displayed menu and clicking on Remove. Now add the variation map from the Available maps menu (by selecting the map and clicking on Add). Make the Variation map as the master map by selecting it and clicking the Make Master/Move to Bottom option. Then click on Apply. Now two maps are displayed, Variation (it's the rightmost and the master map) and Gene\_seq. The master map provides detailed information for the map features, in this case SNPs.". (The Mini-Course Map Viewer Quick Start describes the usage of the Map Viewer in detail.) Zoom in on the blast hit area (red bar). There are two SNPs in the area, one of them is rs334. Click on the link for the SNP. There is an A/T SNP at the nucleotide position 4035473 on the contig NT\_009237 as mentioned under Fasta sequence and Integrated maps. Is this the same nucleotide variation found in the BLAST result in Step 1?

### *Step 4. Determine whether the mutant HBB gene causes a phenotype:*

Go back to the Map Viewer report. Make the Gene\_seq map as the master map. Select the link to the OMIM database. It takes us to the OMIM report for the HBB gene that details how mutations in the HBB gene are associated with a phenotype, sickle cell anemia. As mentioned in the report, the allelic variants are listed for the mature HBB protein which lacks initiator methionine. Click on the Allelic Variant "View list" to get information about mutant proteins from patients. Is Glu6Val variant mentioned in the list? Which phenotype does it cause?

## Summary:

This mini-course describes steps to identify the gene expressing the ESTs obtained from a sickle cell anemia patient, download the gene sequence, identify known SNPs in the gene and find SNP-associated phenotypes.

Step 1: The query EST sequence was found to align to contig NT\_009237.17 on chromosome 11 with one nucleotide difference (T to A with respect to the nucleotide 4035473 on the contig).

Step 2: The query EST was found to be expressed by the HBB gene.

Step 3: The query EST sequence contains a known SNP (T/A with respect to the nucleotide 4035473 on contig NT\_009237.17).

Step 4: Mutations in the HBB gene are associated with sickle cell anemia.